

Reply to the National Research Council Study on Parapsychology¹

JOHN A. PALMER, CHARLES HONORTON, AND JESSICA UTTS

ABSTRACT: In December, 1987, the National Research Council (NRC) of the National Academy of Sciences released its report, *Enhancing Human Performance* (Druckman & Swets, 1988), the result of a study commissioned by the U.S. Army to assess various techniques of improving human performance. Six areas were addressed, one of which was parapsychology. The Committee concluded that it could find "no scientific justification from research conducted over a period of 130 years for the existence of parapsychological phenomena" (p. 22). After a careful review of the NRC report, we have found that although it is written in scientific language, it does not represent an unbiased scientific assessment of parapsychology and that the conclusion of the report is unwarranted. The grounds for this conclusion are presented in this reply.

OVERVIEW

On December 3, 1987, the National Research Council (NRC) of the National Academy of Sciences convened a well-attended press conference in Washington, DC, to announce the release of its report, *Enhancing Human Performance* (Druckman & Swets, 1988). The report is the result of a study commissioned by the U.S. Army to assess various techniques of enhancing human performance. One of the six areas addressed in the report is parapsychology (research involving anomalous communication processes such as extrasensory perception and psychokinesis, collectively known as psi phenomena). At the NRC press conference, John A. Swets, the Committee Chair, stated that "perhaps our strongest conclusions are in the area of parapsychology." Indeed, the report concludes: "The Committee finds no scientific justification from research conducted over a period of 130 years for the existence of parapsychological phenomena" (p. 22).

After carefully reviewing the NRC report, we have found that although it is couched in scientific language, it does not represent an unbiased scientific assessment of parapsychology and that the above conclusion is totally unwarranted. In particular,

¹ Copyright © 1988 by the Parapsychological Association, Inc. Research Triangle Park, NC. This is a special report prepared for the Board of Directors of the Parapsychological Association, Inc. It has been edited in accordance with the style and format of this *Journal* and that of the American Psychological Association, with permission of the authors and the Board of the Parapsychological Association.

1. The single chapter of the report on parapsychology is restricted to four selected areas of research conducted during the past 20 years. The scope of the review is limited to less than 10% of the systematic scientific effort in parapsychology, which began with the Duke University work of J. B. Rhine in the 1930s, and no explanation is provided concerning the purported "130 year" history of research in parapsychology.

2. The two principal evaluators of parapsychological research for the Committee, Ray Hyman and James Alcock, were publicly committed to a negative position on parapsychology at the time the Committee was formed. Both are members of the Executive Council of an organization well known for its zealous crusade against parapsychology. Yet no attempt was made to balance the Committee with scientists who have taken a more positive or a neutral position on parapsychology.

3. Even within this limited scope of review, the Committee's method of assessing parapsychology violates its own stated guidelines for research evaluation, which specify the identification and assessment of plausible alternatives. With regard to the better parapsychological experiments, the Committee admits, "We do not have a smoking gun, nor have we demonstrated a plausible alternative" (p. 200).

4. The report selectively omits important findings favorable to parapsychology contained in one of the background papers commissioned for the Committee, while liberally citing from other papers supportive of the Committee's position. The principal author of the favorable paper, an eminent Harvard psychologist, was actually asked by the Chair of the NRC Committee to withdraw his favorable conclusions.

Obviously the Committee's conclusion far outstrips the scope of its investigation. What it accomplishes is to suggest that parapsychology is a field that has had its day and failed to deliver the goods, which in turn implies that future research is not likely to be fruitful. It thus has the effect of discouraging future research and the funding of such research. The Committee's overall conclusion, as well as many other statements in the report, encourage the reader to ignore the fact that parapsychologists have accumulated a large body of experimental findings that (a) suggest important new means of human interaction with their environment and (b) cannot be plausibly attributed to known conventional mechanisms.

ANATOMY OF PREJUDGMENT: COMPOSITION OF THE NRC PARAPSYCHOLOGY SUBCOMMITTEE

Belief in paranormal phenomena is still growing, and the dangers to our society are real. . . . [I]n these days of government budget-cutting the Defense Department may be spending millions of tax dollars on developing "psychic arms. . . ." Please help us in this battle against the irrational. Your contribution, in any amount, will help us grow and be better able to combat the flood of belief in the paranormal.

The above excerpt is from a fund-raising letter from the Committee for the

Reply to NRC Study

Scientific Investigation of Claims of the Paranormal (CSICOP), dated March 23, 1985, and co-signed by Ray Hyman, Chair of the NRC Subcommittee on Parapsychological Techniques.

In a background paper solicited by the NRC Committee, Griffin (1988) outlines the difficulties in trying to evaluate evidence objectively when one is committed to a belief system. He also emphasizes the strong desire to protect those beliefs.

Probably the most powerful force motivating our desire to protect our beliefs—from others' attacks, from our own questioning, and from the challenge of new evidence—is commitment . . . people will often react to disconfirming evidence by strengthening their beliefs and creating more consonant explanations. This drive to avoid dissonance is especially strong when the belief has led to public commitment. (p. 33)

Although it was Griffin's intent to show how parapsychologists' beliefs can lead them to ignore certain evidence, we contend that the NRC report exemplifies the Committee's need to protect *their* beliefs. Both Hyman, Chair of the NRC Parapsychology Subcommittee, and Alcock, author of the only paper specifically on parapsychological techniques to be commissioned by the Committee (Alcock, 1988), belong to CSICOP's Executive Council and are among its most active members. Hyman has been the Chair of CSICOP's Parapsychology Subcommittee since its inception.

CSICOP is well known for its efforts to debunk parapsychology. It was founded in 1976 by philosopher Paul Kurtz and sociologist Marcello Truzzi, when "Kurtz became convinced that the time was ripe for a more active crusade against parapsychology and other pseudo-sciences" (Pinch & Collins, 1984, p. 527). Truzzi resigned in 1977 "because of what he saw as the growing danger of the committee's excessive negative zeal at the expense of responsible scholarship" (Collins & Pinch, 1982, p. 42). He has since stated that "CSICOP has proclaimed its major aim to be inquiry while actually being centrally concerned with advocacy (i.e., discrediting claims of the paranormal)" (Truzzi, 1982, p. 4). In their own literature, CSICOP makes clear their belief that claims for paranormal phenomena are unreasonable: "Why the sudden explosion of interest, *even among some otherwise sensible people*, [italics added] in all sorts of paranormal 'happenings'?" (CSICOP brochure).

Both Hyman and Alcock were members of the CSICOP Executive Council in 1985, when the NRC Committee was formed, and there was abundant evidence at that time that both men had publicly committed themselves to the belief that scientists who were convinced by the evidence for psychic phenomena were fooling themselves. For example, Hyman (1985a) wrote:

The total accumulation of 130 years' worth of psychical investigation has not produced any consistent evidence for parormality that can withstand acceptable scientific scrutiny. What should be interesting for the scientific establishment is not that there is a case to be made for psychic phenomena

but, rather, that the majority of scientists who decided to seriously investigate *believed* that they had made such a case. How can it be that so many outstanding scientists, including several Nobel Prize winners, have convinced themselves that they have obtained solid scientific evidence for paranormal phenomena? (p. 7)

Alcock (1981) expressed the same view somewhat more colorfully:

Parapsychology is indistinguishable from pseudo-science, and its ideas are essentially those of magic. This does not of course mean that psi does not exist, for one cannot demonstrate the non-existence of psi any more than one can prove the non-existence of Santa Claus. But let there be no mistake about the empirical evidence: There is *no* evidence that would lead the cautious observer to believe that parapsychologists and paraphysicists are on the track of a real phenomenon, a real energy or power that has so far escaped the attention of those people engaged in "normal" science. There is considerable reason, on the other hand, to believe that human desires and human self-delusion are responsible for the durability of parapsychology as a formal endeavour." (p. 196)

Given the attitudes of these two individuals, the biased nature of the NRC report was easily predictable. The Army presumably expected the National Research Council to provide sound and unbiased advice regarding this controversial subject. Why then was the Parapsychology Subcommittee composed the way it was? Almost half a million dollars of taxpayers' money was spent on the NRC report. These taxpayers, as well as the scientific community and the Army itself, deserve an answer to this question.

EVALUATION ISSUES

Describing the Committee's standards for evaluating basic research, Druckman and Swets (1988) write:

Ruling out *alternative explanations or mechanisms* [italics added] requires intimate knowledge of a research area. Historical findings and critical commentary are needed to identify alternatives, *determine their plausibility* [italics added], and judge how well they have been ruled out in particular sets of experiments. (p. 25)

With respect to parapsychology, the Committee admits that "we do not have a smoking gun, nor have we demonstrated a plausible alternative" (p. 200). They go on, however, to suggest that it is not necessary for them to do so in order to justify their wholesale dismissal of the evidence. What they offer instead is the metaphor of a "dirty test tube," which they describe as follows:

The critic does not claim that the results have been produced by some artifact [italics added], but instead points out that the results have been obtained under conditions that fail to meet generally accepted standards. The gist of this type of criticism is that test tubes should be clean when doing careful

and important scientific research. To the extent that the test tubes were dirty, it is suggested that the experiment was not carried out according to acceptable standards. Consequently, *the results remain suspect even though the critic cannot demonstrate that the dirt in the test tubes was sufficient to have produced the outcome* [italics added]. Hyman's critique of the Ganzfeld psi research and Alcock's [background] paper on remote viewing and random number generator research are examples of this type of criticism. (pp. 199-200)

This approach directly contradicts the Committee's own guidelines for the evaluation of scientific evidence, as illustrated by the above quote. Argument by metaphor is no substitute for systematic scientific analysis, and no justification is provided for this singular violation of the Committee's own standards.

It is clear that the Committee's failure to identify plausible alternative explanations for many of the parapsychological studies it reviews is not for want of trying. For example, Hyman attempted to demonstrate a statistical relationship between putative methodological flaws and study outcomes in the Ganzfeld experiments. In his initial analysis, Hyman claimed an almost perfect linear correlation between his flaw assignments and the study outcomes. This analysis contained a large number of errors that Hyman later attributed to typing errors (personal communication to Honorton, November 29, 1982). His next published critique (Hyman, 1985b) was based on a complex multivariate analysis that was subsequently shown to be meaningless (Saunders, 1985). Finally, Hyman agreed that "the present data base does not support any firm conclusion about the relationship between flaws and study outcome" (Hyman & Honorton, 1986, p. 353).

The Committee's approach to evaluating evidence for parapsychological processes is scientifically sterile. Its resort to the "dirty test tube" metaphor provides an unrestricted license for the wholesale dismissal of research findings on the basis of vague and ad hoc "weaknesses." Whole domains of research are dismissed through allusions to "inadequate documentation," "inadequate controls," "overcomplicated experimental setups," and "lack of experimental rigor." Yet if the measure of good scientific methodology is its capacity to rule out plausible alternatives, such attributions are clearly inappropriate. No scientific experiment is so pristine that it can withstand the efforts of a sufficiently determined critic, but the fact remains that by any reasonable standard the methodology of many successful psi experiments is fundamentally sound.

TREATMENT OF THE SCIENTIFIC EVIDENCE

Despite the impression given by the Committee's overall conclusion, quoted above, their assessment of the scientific evidence for parapsychological processes is restricted to a few selected areas of research. In this

section, we respond to the Committee's treatment of research in these selected areas. Two of the research domains considered, the psi Ganzfeld and random number generator (RNG) experiments, have been the subject of recent meta-analytic investigations (Honorton, 1985; Hyman, 1985b; Radin & Nelson, 1987), and like a number of other psi research areas that are ignored by the Committee, these domains have been found to have replication rates comparable to those in other areas of psychology. Because the Committee members are themselves psychologists, we were surprised by their characterization of replicability in parapsychology:

The type of replicability that has been claimed so far is the possibility of obtaining significant departures from the chance baseline *in only a proportion of the experiments* [italics added], which is a kind of replicability quite different from the consistent and lawful patterns of covariation found in other areas of inquiry. (pp. 174-175)

Contrary to the Committee's statement, less than perfect replicability is the rule and not the exception in most areas of the behavioral sciences, and this is one of the reasons why meta-analytic techniques for assessing whole areas of behavioral research have become so widely used. It is not uncommon in psychology for conclusions to be drawn from studies in which replications have not even been attempted. Replicability problems have been widely acknowledged in recent years by experts in areas as diverse as the neurochemistry of learning and memory (Dunn, 1980) and medical studies of placebo efficacy (Moermaonn, 1981). Even such "hard-science" areas as laser construction have sometimes been beset by replication problems (Collins, 1974).

As for "consistent and lawful patterns of covariation," the limited scope of the Committee's review, which focuses exclusively on evidence for the *existence* of parapsychological effects in a few selected areas, completely ignores parapsychological research oriented toward identification of patterns of covariation of psi performance with other variables. There is, however, a considerable body of such research. Although a detailed discussion of this work is beyond the scope of the present paper, three such areas may be briefly mentioned:

Extraversion/Introversion. Beginning in the early 1940s numerous attempts have been made to correlate experimental ESP performance with individual differences in subjects' personality and attitudinal characteristics. Palmer (1977), in a review of 33 experiments involving the relationship between ESP performance and standard psychometric measures of extraversion/introversion, found that extraverts scored higher than introverts in 70% of these experiments ($p = .017$) and all eight of the significant relationships showed superior ESP performance by extraverts ($p = .0039$).

Belief in ESP. In a review of 17 experiments testing the hypothesis that subjects who believed in ESP would show superior ESP performance compared to subjects who did not believe in ESP, Palmer (1971) found

that the predicted pattern occurred in 76% of the experiments ($p = .024$), and all six of the experiments with individually significant outcomes were in the predicted direction ($p = .015$).

Hypnosis. Schechter (1984) reported a meta-analysis of studies comparing the effect of hypnotic induction and nonhypnosis control procedures on performance in ESP card-guessing tasks. There were 25 experiments by investigators in 10 different laboratories. Consistently superior ESP performance was found to occur in the hypnotic induction conditions compared to the control conditions of these experiments ($p = .006$).

Detailed evaluations of these and other areas involving the systematic covariation of psi performance with other variables are presented elsewhere (e.g., Eysenck, 1967; Honorton, 1977; Johnson & Haraldsson, 1984; Palmer, 1985; Schmeidler & McConnell, 1958/1973; Stanford, 1987).

Psi Ganzfeld Research

ESP research using Ganzfeld stimulation, a mild form of perceptual isolation, has its origin in earlier research linking psi effects to dreaming, hypnosis, and other internal attention states characterized by functional sensory deprivation. A typical Ganzfeld experiment involves a sender and a receiver, each sequestered in separate, acoustically isolated rooms. The receiver, undergoing perceptual isolation, attempts to describe a randomly selected target picture presented to the physically remote sender by providing a continuous verbal report of ongoing imagery and associations. Upon completion of the session, the receiver, on a blind basis, attempts to identify the actual target from a judging pool containing the target and three or more control pictures.

A meta-analysis of 28 psi Ganzfeld studies by investigators in 10 different laboratories (Honorton, 1985) found a combined z score of 6.6, a result associated with a probability of less than 1 part in a billion. Independently significant outcomes have been reported by 6 of the 10 investigators, and the overall significance is not dependent on the work of any one or two investigators. Moreover, in order to account for the observed experimental results on the basis of selective reporting, it would be necessary to assume that there were more than 400 unreported studies averaging chance results.

The NRC Committee's assessment of ESP Ganzfeld research, which is based on Hyman's (1985b) critique, clearly illustrates its systematic failure to cite information unfavorable to its case in evaluating parapsychology. For example, although Hyman has subsequently conceded "that there is an overall significant effect in [the ESP Ganzfeld] data base that cannot reasonably be explained by selective reporting or multiple analysis" (Hyman & Honorton, 1986, p. 351) and that "significant outcomes have been produced by a number of different investigators" (p. 352), neither of these important points is mentioned anywhere in the Committee's report.

The Committee drew heavily on Palmer's (1985) report to the Army, which includes reviews of the major research projects it evaluated. Palmer made several criticisms of these projects and the Committee cites these favorably. However, Palmer also made several criticisms of points made by other critics of the experiments (including a critique of Hyman's flaw analyses of the Ganzfeld research); these criticisms were consistently ignored.

A particularly revealing example of the Committee's selective bias is its treatment of a background paper the Committee solicited from Monica Harris and Robert Rosenthal of Harvard University (Harris & Rosenthal, 1988a). Rosenthal is a leading social science methodologist and a pioneer in the development of meta-analytic techniques for evaluating entire research domains. Harris and Rosenthal have no prior involvement with parapsychology, nor have they taken a public position on this controversial subject. They undertook a comparative study of the major topics reviewed by the Committee and concluded that "only the Ganzfeld ESP studies [the only psi studies they evaluated] regularly meet the basic requirements of sound experimental design" (p. 53). On a 25-point scale of "overall quality," the Ganzfeld experiments were given a rating of 19, whereas the other (nonparapsychological) areas reviewed received ratings from 3 to 13. This relative weighting of research quality is diametrically opposed to that of the Committee.

In its evaluation of parapsychology, the Parapsychology Subcommittee extensively cites the background papers of Alcock (1988) and Griffin (1988), with whose conclusions it agrees, but it does not cite the Harris and Rosenthal paper even once. (Their paper is, however, cited elsewhere in the report, in the evaluation of other, nonparapsychological areas by the Subcommittee on Accelerated Learning.) Incredibly, at one stage of the process, John Swets, Chair of the Committee, actually phoned Rosenthal and asked him to withdraw the parapsychology section of his paper. When Rosenthal declined, Swets and Druckman then requested that Rosenthal respond to criticisms that Hyman had included in a July 30, 1987 letter to Rosenthal.

If the Committee considered Hyman's critique of the Harris-Rosenthal paper to be effective, why did they not simply challenge the latter in the report itself? The overall behavior of the Committee suggests that it was eager to give Rosenthal's dissenting views as little exposure as possible.

In response to Hyman's letter, Harris and Rosenthal prepared a postscript to their paper. Rosenthal conducted additional analyses of the putative relationship between psi Ganzfeld study outcomes and potential methodological flaws:

The heart of the matter is the relationship of flaws to research results and that is what our analyses are designed to investigate. In a 1986 manuscript Hyman suggested that the relationship of flaws to study outcomes should be examined in a multivariate manner. Accordingly, that is the nature of our analyses in our first pass effort to examine the likelihood that methodolog-

ical flaws are driving the results of the Ganzfeld studies to an appreciable degree. (Harris & Rosenthal, 1988b, pp. 1-2)

Using Hyman's own most recent flaw ratings, Rosenthal failed to find any significant relationships between flaws and ESP significance levels and effect sizes in each of two multivariate analyses. Harris and Rosenthal concluded:

Our analysis of the effects of flaws on study outcome lends no support to the hypothesis that Ganzfeld research results are a significant function of the set of flaw variables. In addition, a series of 10 new studies designed to control for earlier presumed flaws yielded results quite consistent with the original set of 28 studies. (Harris & Rosenthal, 1988b, p. 3)

The Committee's response to the Harris and Rosenthal postscript became available to us after the completion of this report. Our comments on the Committee's response are presented in the Appendix.

Random Number Generator (RNG) Experiments

Psi research with random number generators (RNGs) involves attempts by subjects to introduce a bias in the normally random output of an electronic device, solely by intention. The randomness of the device is typically based on radioactive decay, electronic diode noise, or randomly seeded pseudorandom sources. Nearly all of these experiments involve a variety of controls against experimental error. These include automated data recording, trial-by-trial oscillation of the target definition, alternation of subjects' attempts to produce high or low scores, and control runs in which large numbers of RNG trials are collected without attempts to influence the device. A comprehensive meta-analytic review of the RNG research literature encompassing all known RNG studies between 1959 and 1987 has been reported by Radin and Nelson (1987), comprising over 800 experimental and control studies conducted by a total of 68 different investigators. The overall *z* score for the 597 experimental series was 15.76 ($p < 10^{-35}$), whereas 235 control series yielded an overall *z* score of -0.67, a result well within the range of chance fluctuation. In order to account for the observed experimental results on the basis of selective reporting, it would be necessary to assume that there were more than 50,000 unreported studies averaging chance results.

The NRC Committee concedes that the RNG research cannot reasonably be explained by chance (p. 207). Although it criticizes the methods used to test the randomness of these devices under control conditions, the Committee admits that "the critics have not specified any plausible mechanisms that would account for the obtained differences between the experimental and control trials" (p. 187).

In contrast to the Radin and Nelson meta-analysis, the Committee's discussion of the RNG research is limited exclusively to the work of two

investigators, Helmut Schmidt of the Mind Science Foundation in San Antonio and Robert Jahn of the Princeton University School of Engineering. Although the Committee is correct in stating that Schmidt and Jahn have been the two largest contributors to the RNG research literature, the overall significance of this research area is not dependent on the contribution of these two investigators. When the work of Schmidt and Jahn is removed, the overall z score for the remaining 66 investigators is still 5.48, a result that is associated with a probability of less than 1 in 40 million (Radin, May, & Thomson, 1985, p. 216).

Radin and Nelson evaluated the impact of potential flaws on study outcomes by assigning to each experiment a single quality weight based upon 16 criteria derived in part from published criticisms of RNG studies by major critics including Hyman and Alcock. These quality criteria fell into four categories: RNG integrity, data integrity, statistical integrity, and procedural integrity. No significant relationship was found between research quality and study outcome. Moreover, the unweighted and quality-weighted effect sizes were nearly identical.

The Committee refers to an unnamed physicist "who claims to have several years of experience in constructing and testing random number devices" who told them that "it is quite possible, under some circumstances, for the human body to act as an antenna and, as a result, possibly bias the output" (p. 190). Inexplicably, the Committee makes no attempt to document this claim, which if true, could conceivably provide a plausible alternative to *some* RNG study outcomes. In view of the Committee's central mission for the Army, it is odd that no attempt was made to follow this up, because the military's interest in RNG effects is related to human interaction with delicate electronics. If this unnamed source's claim is valid, it could have important implications, irrespective of the "paranormality" of the effect. Yet the Committee was content to simply repeat this undocumented assertion for the purpose of casting doubt on the RNG research. Nor does the Committee acknowledge the fact that the purported antenna effect, even if it were shown to be valid in some instances, is totally irrelevant to many of the RNG studies (including studies by Schmidt and Jahn), such as those using prerecorded targets or pseudo-random sources and studies in which the target definition is oscillated on a trial-by-trial basis.

The Committee favorably cites one RNG experiment by May, Humphrey, and Hubbard (1980b), but they downgrade its importance because of its "marginal" level of significance ($p = .029$). However, this criticism is completely spurious. The May et al. experiment used a standard sequential sampling technique (Wald, 1947), which was designed not to achieve a highly significant p value in a fixed number of trials but to reach a preset significance level ($p \leq .05$) in as few trials as possible and then stop. Sequential sampling was a primary feature of the May et al. experiment, and it is inconceivable that this fact could be overlooked by anyone reading their report. In order to avoid problems involving optional

stopping, the analysis of this experiment, which was specified in advance (May, Humphrey, & Hubbard, 1980a), was based on the number of subjects who successfully met the sequential sampling criteria, and the reported significance level represents the probability that two or more subjects would meet these criteria. Although the Committee complains that "only two of the seven subjects produced significant results" (p. 189), they fail to mention that this success rate is nearly six times that expected by chance.

The Committee fails to cite a number of other experiments in which their criticisms are demonstrably inapplicable. A prime example is the recent experiment of Schmidt, Morris, and Rudolph (1986), the protocol for which was published prior to the beginning of the experiment (Schmidt, Morris, & Rudolph, 1982). An important control feature of this experiment, which used prerecorded targets, is that responsibility for the test was shared by three different investigators in two independent laboratories. The Committee's failure to mention this experiment is especially surprising because Alcock (1988), who reviewed the experimental report for the Committee and even had access to the pertinent raw data, admitted that he could find nothing seriously wrong with it.

The Committee alludes to the possibility of data tampering by subjects in RNG experiments, but it offers no suggestions as to how such tampering could have been accomplished. Subject fraud has never been demonstrated in RNG research, and we are aware of no evidence that would justify suspicion of fraud in any of the experiments under review. Moreover, one of the 16 quality criteria employed in the Radin/Nelson meta-analysis was "Unselected Subjects." Studies using the experimenter, self-proclaimed "psychics," or otherwise special subjects were penalized. No significant difference was found between studies using unselected and selected subjects.

Backster's "Primary Perception"

Many readers of the Committee's report who are not familiar with the technical literature of parapsychology might be impressed by the fact that the Committee's critique of parapsychology is based not only on its examination of selected portions of the research literature, but also upon "experimental work that the committee *actually witnessed* [italics added] by visiting a parapsychological laboratory" (p. 169). Statements of this sort occur in several places in the report. In fact, the experiment witnessed by the Committee was not that of a parapsychologist, nor did it take place in a "parapsychological laboratory." The experiment was that of a polygraph specialist, Cleve Backster, and it took place at the Backster Research Foundation. Neither Backster nor his institution has ever had any affiliation with scientific parapsychology. Because all attempts by serious investigators to replicate Backster's initial claim have been unsuccessful, Back-

ster's claims are not taken seriously within mainstream parapsychology. Nevertheless, the Committee's characterization, quoted above, encourages readers to assume that Backster's work is representative of parapsychological research in general. This assumption is strongly reinforced by the fact that the Committee devotes nearly as much space to Backster as it does to the RNG research and the Backster claim receives almost twice as much space as the Ganzfeld research domain. Even though the Committee acknowledges that Backster's research is "at a far less developed stage" (p. 193), the fact that it is treated in the same format as this other research and is given so much attention creates the impression that it is representative of parapsychological research in general. It most certainly is not. Backster's research illustrates a lack of sensitivity to the needs of proper experimental control that exceeds even the most problematic of the research considered by the Committee and would never be accepted for publication in a mainstream parapsychological journal. It does, however, illustrate research in which a plausible, empirically grounded alternative explanation does exist and a negative conclusion is therefore justified.

POTENTIAL APPLICATIONS

Because the NRC report was funded by the Army for the express purpose of assessing potential applications, the Committee has been remiss in ignoring controlled psi studies directed toward eventual applications (e.g., Ryzl, 1966; Carpenter, 1975; Puthoff, 1985; Puthoff, May, & Thomson, 1986).

These experiments all used statistical averaging techniques based on information theory to enhance accuracy. In Ryzl's experiment, for example, a subject was asked to guess on repeated trials whether the green or white face of each of a set of concealed cards was uppermost. Although the subject's rate of success on individual calls was only 62%, translation of the "majority vote" for each card resulted in the identification, without error, of 15 binary-encoded decimal digits. Carpenter's experiment resulted in the successful identification of the binary (Morse) code equivalent to the word "PEACE." A successful effort to improve the accuracy of detection of random binary sequences generated by hidden roulette-wheel spins (red/black) and coin tosses (heads/tails) was reported by Puthoff (1985). Initial success rates of 52.6%, 55%, and 60% were amplified, respectively, to 60%, 60%, and 71%. Later, Puthoff et al. (1986) reported encouraging results based on a more efficient statistical averaging procedure.

Although no responsible parapsychologist would claim that reliable practical applications of psi abilities have yet been achieved, more intensive applications-oriented basic research is clearly justified by the success of these and similar efforts. Instead of seriously addressing the challenge

Reply to NRC Study

posed by such research, the Committee ignores this work completely and expresses its negative bias against potential psi applications through reference to "warrior monks" and "hyperspatial howitzers" (p. 171). Clearly, the Committee's mission for the Army has not been well served by this type of frivolousness.

QUALITATIVE EVIDENCE AND SUBJECTIVE BIAS

The section entitled "The Problem of Qualitative Evidence" (pp. 200-206) vividly demonstrates the very "problem" it condemns. Perhaps aware of the weakness of its case against the scientific evidence, the Committee tried to strengthen it by insinuating that parapsychologists have been biased by personal psychic experiences. The clear implication is that they cannot be trusted to carry out basic experimental procedures in a competent fashion. Highly prejudicial statements of the following sort riddle these pages: "More typical is the proponent who, as a result of compelling personal experience, not only has no doubt about the reality of underlying paranormal cause, but also has no patience with the refusal of others to support that belief" (p. 202). The whole purpose of this section, presumably, is to argue that conclusions of fact should be based upon empirical evidence obtained by proper scientific methods. Yet no such evidence is offered to support the assertion that attitudes such as those reflected in the above quote are shared by even a significant minority of parapsychologists.

Obviously the interpretation of profound personal experiences is open to all sorts of potential biases and distortions. Clearly such experiences are no substitute for scientific evidence. On the other hand, when careful consideration of such alternative explanations renders them implausible, it is not unreasonable to entertain the possibility of psi, unless, of course, one "knows" in advance that psi events are impossible. It does not follow that having a personal psychic experience disqualifies one from conducting rigorous scientific research any more than a deep appreciation for the grandeur of the universe disqualifies one from doing sound research in astrophysics. Scientists frequently choose to study specific research topics out of curiosity generated by events in their personal lives.

CONCLUSIONS

In this response we have documented some, but by no means all, of the problems with the Committee's report. As we have seen, the Committee's primary conclusion regarding parapsychology is not merely unjustified by their report, it is directly contradicted by the Committee's admission that it can offer no plausible alternatives. This concession, coming as it does from a Committee whose principal evaluators of parapsychology were publicly committed to a negative verdict at the outset of their investiga-

tion, actually constitutes a strong source of support for the conclusion that parapsychology has identified genuine scientific anomalies.

We have documented numerous instances where, in lieu of plausible alternatives, the Committee's attempts to portray parapsychology as "bad science" have been based upon erroneous or incomplete descriptions of the research in question, rhetorical enumeration of alleged "flaws" that by its own admission frequently have no demonstrable empirical consequences, selective reporting of evidence favorable to its case, and the selective omission of evidence not favorable to its case. Moreover, with respect to the Committee's central mission for the U.S. Army, we have shown that the Committee's prejudice against parapsychology has led it to ignore research, the further development of which could have important implications for our national security.

The scientific and defense communities are entitled to a rigorous and unbiased assessment of this research area. A strong *prima facie* case has been made for the existence of psi anomalies, and meaningful relationships between such events and psychological variables have been reported in the literature. Further efforts and resources should be expended toward the identification of underlying mechanisms and the development of theoretical models, either conventional or "paranormal," that can provide adequate understanding.

REFERENCES

ALCOCK, J. E. (1981). *Parapsychology: Science or Magic?* New York: Pergamon Press.

ALCOCK, J. E. (1988). *A Comprehensive Review of Major Empirical Studies in Parapsychology Involving Random Event Generators or Remote Viewing.* Washington, DC: National Academy Press.

CARPENTER, J. C. (1975, January). *Toward the effective utilization of enhanced weak-signal ESP effects.* Paper presented at the Meeting of the American Association for the Advancement of Science, New York, NY.

COLLINS, H. M. (1974). The TEA set: Tacit knowledge and scientific networks. *Science Studies*, 4, 165-186.

COLLINS, H. M., & PINCH, T. J. (1982). *Frames of Meaning: The Social Construction of Extraordinary Science.* Boston: Routledge & Kegan Paul.

DRUCKMAN, D., & SWETS, J. A. (Eds.). (1988). *Enhancing Human Performance: Issues, Theories, and Techniques.* Washington, DC: National Academy Press.

DUNN, A. J. (1980). Neurochemistry of learning and memory: An evaluation of recent data. *Annual Review of Psychology*, 31, 343-390.

EYSENCK, H. J. (1967). Personality and extra-sensory perception. *Journal of the Society for Psychical Research*, 44, 55-71.

Reply to NRC Study

GRiffin, D. (1988). *Intuitive Judgment and the Evaluation of Evidence.* Washington, DC: National Academy Press.

HARRIS, M. J., & ROSENTHAL, R. (1988a). *Interpersonal Expectancy Effects and Human Performance Research.* Washington, DC: National Academy Press.

HARRIS, M. J., & ROSENTHAL, R. (1988b). *Postscript to Interpersonal Expectancy Effects and Human Performance Research.* Washington, DC: National Academy Press.

HONORTON, C. (1977). Psi and internal attention states. In B. B. Wolman (Ed.), *Handbook of Parapsychology* (pp. 435-472). New York: Van Nostrand Reinhold.

HONORTON, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology*, 49, 51-91.

HYMAN, R. (1985a). A critical historical overview of parapsychology. In P. Kurtz (Ed.), *A Skeptic's Handbook of Parapsychology* (pp. 3-96). Buffalo, NY: Prometheus Books.

HYMAN, R. (1985b). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology*, 49, 3-49.

HYMAN, R. (1988). *Comments on the Postscript to Harris and Rosenthal.* Washington, DC: National Academy Press.

HYMAN, R., & HONORTON, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology*, 50, 351-364.

JOHNSON, M., & HARALDSSON, E. (1984). The defense mechanism test as a predictor of ESP scores: Icelandic studies IV and V. *Journal of Parapsychology*, 48, 185-200.

MAY, E. C., HUMPHREY, B. S., & HUBBARD, G. S. (1980a). *Phase I: Hardware Construction and System Evaluation* (SRI Project 8585). Menlo Park, CA: SRI International.

MAY, E. C., HUMPHREY, B. S., & HUBBARD, G. S. (1980b). *Electronic System Perturbation Techniques* (Final Report). Menlo Park, CA: SRI International.

MOERMANN, D. E. (1981). Edible symbols: The effectiveness of placebos. In T. A. Seboek & R. Rosenthal (Eds.), *The Clever Hans Phenomenon* (pp. 256-258) [Annals of the New York Academy of Sciences, Vol. 364]. New York: New York Academy of Sciences.

PALMER, J. [A.] (1971). Scoring in ESP tests as a function of belief in ESP. Part I: The sheep-goat effect. *Journal of the American Society for Psychical Research*, 65, 373-408.

PALMER, J. A. (1977). Attitudes and personality traits in experimental ESP research. In B. B. Wolman (Ed.), *Handbook of Parapsychology* (pp. 175-201). New York: Van Nostrand Reinhold.

PALMER, J. A. (1985). *An Evaluative Report on the Current Status of Parapsychology.* Contract DAJA 45-84-M-0405. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

PINCH, T. J., & COLLINS, H. M. (1984). Private science and public knowledge: The Committee for the Scientific Investigation of Claims of

the Paranormal and its use of the literature. *Social Studies of Science*, 14, 521-546.

PUTHOFF, H. E. (1985). Calculator-assisted psi amplification. In R. A. White & J. Solfvin (Eds.), *Research in Parapsychology 1984* (pp. 48-51). Metuchen, NJ: Scarecrow Press.

PUTHOFF, H. E., MAY, E. C., & THOMSON, M. J. (1986). Calculator-assisted psi amplification II: Use of the sequential-sampling technique as a variable-length majority-vote code. In D. H. Weiner & D. I. Radin (Eds.), *Research in Parapsychology 1985* (pp. 73-76). Metuchen, NJ: Scarecrow Press.

RADIN, D. I., MAY, E. C., & THOMSON, M. J. (1985). Psi experiments with random number generators: Meta-analysis Part 1. In *Proceedings of Presented Papers: The Parapsychological Association 28th Annual Convention*, Vol. 1, 201-233.

RADIN, D. I., & NELSON, R. D. (1987). *Replication in Random Event Generator Experiments: A Meta-Analysis and Quality Assessment* (Technical Report 87001). Princeton, NJ: Princeton University Human Information Processing Group.

RYZL, M. (1966). A model of parapsychological communication. *Journal of Parapsychology*, 30, 18-30.

SAUNDERS, D. R. (1985). On Hyman's factor analyses [Appendix B in Honorton, 1985]. *Journal of Parapsychology*, 49, 86-88.

SCHECHTER, E. I. (1984). Hypnotic induction vs. control conditions: Illustrating an approach to the evaluation of replicability in parapsychological data. *Journal of the American Society for Psychical Research*, 78, 1-27.

SCHMEIDLER, G. R., & MCCONNELL, R. A. (1973). *ESP and Personality Patterns*. Westport, CT: Greenwood Press. (Original work published 1958)

SCHMIDT, H., MORRIS, R. L., & RUDOLPH, L. (1982). Channeling psi evidence to critical observers. In W. G. Roll, R. L. Morris, & R. A. White (Eds.), *Research in Parapsychology 1981* (pp. 136-138). Metuchen, NJ: Scarecrow Press.

SCHMIDT, H., MORRIS, R., & RUDOLPH, L. (1986). Channeling evidence for a PK effect to independent observers. *Journal of Parapsychology*, 50, 1-16.

STANFORD, R. G. (1987). Ganzfeld and hypnotic-induction procedures in ESP research: Toward understanding their success. In S. Krippner (Ed.), *Advances in Parapsychological Research 5* (pp. 39-76). Jefferson, NC: McFarland.

T[RUZZI], M. (1982). Editorial. *Zetetic Scholar*, No. 9, 3-5.

WALD, A. (1947). *Sequential Analysis*. New York: Dover.

Foundation for Research on the Nature of Man
Box 6847, College Station
Durham, North Carolina 27708

Psychophysical Research Laboratories
301 College Road East
Princeton, New Jersey 08544

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

APPENDIX

On the NRC Committee's Response to Harris and Rosenthal's Postscript

Hyman (1988) attempts to justify the NRC Committee's negative position on the Ganzfeld experiments in a reply to Harris and Rosenthal's (1988b) postscript. He argues that the Ganzfeld studies "are not truly independent of one another" (p. 7), presumably because of commonalities in multiple studies contributed by different investigators. The general thrust of Hyman's reply indicates that he now believes that the Ganzfeld research should be analyzed by investigators rather than by studies. He also suggests that Rosenthal's analysis may have lacked sufficient statistical power to demonstrate a significant relationship between putative flaws and study outcome. The gist of Hyman's argument is that no conclusions can be drawn from the Ganzfeld database because it lacks robustness. This claim is demonstrably false.

The robustness argument hinges on a disagreement between Hyman (1985b) and Honorton (1985) in their evaluations of the adequacy of the randomization method employed in eight experiments by Carl Sargent and his colleagues at Cambridge University. Hyman classified these experiments as inadequate on randomization and Honorton classified them as adequate. Hyman writes:

When a data base is so unstable that *just a single change* [italics added] on a disputed point yields a different conclusion, the data base lacks robustness. Statisticians have devised a number of indicators . . . to assess the robustness of a given data base. When such indicators inform the investigator that the alteration or removal of just one case—or even a few cases—can alter the conclusions this warns the investigator against drawing any conclusions. (Hyman, 1988, pp. 2-3)

We agree, but the disagreement between Hyman and Honorton does not, as Hyman implies, involve just a single case. It involves eight cases, or 28% of the experiments in the Ganzfeld database. Although we can find

no legitimate grounds for the wholesale elimination of the Sargent studies, the disposition of Sargent's experiments affects neither the effect size nor the statistical significance of the Ganzfeld database. As Harris and Rosenthal have shown (1988a, Table 4A), when Sargent's studies are removed from the database, the median effect size is unchanged. When the analysis is by investigators, rather than studies, as Hyman now believes is more appropriate, dropping the Sargent studies actually *increases* the median effect size for the remaining nine investigators. The overall *z* score for the 9 remaining investigators is still 5.07 ($p = 2.1 \times 10^{-7}$; that is, one part in 5 million). If, in addition, the studies of Honorton, the other major contributor to the Ganzfeld database, are also eliminated, the overall *z* score for the 8 surviving investigators is still 3.67 ($p = 1.2 \times 10^{-4}$; less than one part in 8 thousand). In other words, 50% of the studies can be removed without jeopardizing the statistical significance of the Ganzfeld database.

Hyman points out that the median *z* score for eight undisputed experiments—those he and Honorton agree involved adequate methods of randomization—is 0.185, which is consistent with chance. But this is misleading. The eight undisputed studies were contributed by five different investigators, two of whom obtained highly significant outcomes. When these undisputed studies are analyzed by investigator, as Hyman now advocates, the overall *z* score is 2.32 ($p = .01$). Thus, the Ganzfeld database remains significant even when more than 70% of the studies are removed.

Hyman's claim that "if we eliminate Sargent's experiments, all the experiments that reported significance were deficient in allowing for the possibility of sensory leakage" (p. 7) is also misleading. Although the statement is true when applied to experiments involving a methodology based on the blind-judging of target sets, it ignores the outcomes of studies using a different methodology based on binary coding of predefined targets. Hyman has previously acknowledged that the binary-coding methodology is not susceptible to potential sensory leakage (Hyman & Honorton, 1986, p. 355). There were five such experiments by three different investigators, (Honorton, 1985, p. 69). Three of the experiments, by two different investigators, were independently significant, and the overall *z* score for all of the binary coding studies was 2.84 ($p = .0023$).

Harris and Rosenthal refer to a new series of experiments by Honorton, designed to control against presumed flaws in some of the earlier Ganzfeld studies. Here we find ourselves in partial agreement with Hyman. Because this new evidence has not yet been fully reported in a peer-reviewed journal, it is understandable that the NRC Committee did not consider it in its report. Although the NRC Committee has been inconsistent in this regard (e.g., its consideration of the as yet unpublished May et al. RNG study), we subscribe to the general principle that potentially important new scientific findings should undergo rigorous peer review before they

are taken seriously. We understand that the new series of Ganzfeld studies by Honorton is currently being prepared for publication.

To summarize, we have shown that (a) the significance of the Ganzfeld database is not dependent upon the studies of any one investigator; (b) analysis by investigators rather than studies, as currently advocated by Hyman, does not appreciably affect the significance of the Ganzfeld database; and (c) even when more than 70% of the studies are removed, the Ganzfeld database remains significant. Under these circumstances, and contrary to Hyman's claim, the Ganzfeld database is very robust.